DOCUMENT RESUME

ED 274 079 EA 018 842

AUTHOR Case, Roland; And Others
TITLE Evaluation for Development Education: An Introduction.
INSTITUTION British Columbia Univ., Vancouver. Center for the Study of Curriculum and Instruction.
REPORT NO ISBN-0-88865-378-6
PUB DATE 85
NOTE 54p.
PUB TYPE Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Educational Administration; *Educational Change; *Educational Planning; Elementary Secondary Education; Evaluation Criteria; *Evaluation Methods; Evaluation Problems; Foreign Countries; *Program Development; *Program Evaluation; Program Improvement
IDENTIFIERS *Canada; *Development Education

ABSTRACT

This handbook outlines a 10-stage approach to conducting evaluations of development education programs. The introducation addresses concerns of the development education community, including the need for a defensible concept of evaluation and for evaluation guidelines for nonprofessional evaluators. Stage 1, "Identifying the Stakeholders," describes the roles of those in a development education program. Stage 2, "Determining the Stake," has three phases: solicitation of program concerns from stakeholders, sorting of concerns, and selection of concerns that will guide evaluation. Stage 3, "Spelling out the Evaluation Mandate," briefly describes the set of instructions given to the evaluator for direction. Stage 4, "Picking the Team," discusses who should conduct the evaluation. Stage 5, "Formulating the Design," outlines four aspects of each evaluation aim, which comprise the evaluation design: standards, evidence, sources, and methodology. Ten tables present data collection measures and other design methods. Stage 6, "Constructing the Instruments," offers recommendations for data-gathering measures. Stage 7, "Collecting the Data," considers the representativeness of results in relation to reporting rate and information reliability. Stage 8, "Tabulation and Interpretation," outlines methods of displaying information and drawing inferences. Stage 9, "Reporting," gives a sample outline of the formal report, and stage 10, "Utilization of Results," recommends inclusion of follow-up provisions. Fifteen references are appended. (CJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

FD274079

EVALUATION FOR DEVELOPMENT EDUCATION

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

LeRoi B. Daniels

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EA 018 842

CASE · WERNER · ONNO · DANIELS

EVALUATION FOR DEVELOPMENT EDUCATION: AN INTRODUCTION

Roland Case
Walter Werner
Eha Onno
LeRoi Daniels

Centre for the Study of Curriculum and Instruction
University of British Columbia
Vancouver

Copyright © 1985
Centre for the Study of Curriculum and Instruction
Faculty of Education
University of British Columbia
Vancouver, British Columbia
V6T 1Z5

All rights reserved. The use of any part of this publication reproduced, transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, or stored in a retrieval system, without the prior consent of the publisher is an infringement of the copyright law.

Cover Design: Michael Howell-Jones

Canadian Cataloguing in Publication Data

Main entry under title:

Evaluation for Development Education

Bibliography: p.
ISBN 0-88865-378-6

1. Evaluation research (Social action programs)
2. Evaluation research (Social action programs) - Case studies.
3. Acculturation - Evaluation.
I. Case, Roland, 1951-
II. University of British Columbia
Centre for the Study of Curriculum and Instruction

H62.E92 1985 303.4'82'072 C85-091225-3

Printed in Canada

ACKNOWLEDGEMENTS

Funding for this handbook was provided by the

Public Participation Program
Special Programs Branch
Canadian International Development Agency

Special thanks is offered to **Moyra Tooke** of **Inter Pares** who was instrumental in launching the project.

Our appreciation goes to **Tim Brodhead** and **Brent Copley** of **The North-South Institute** for their insightful questions and suggestions.

PREFACE

This handbook is intended as a *working document* in two senses of the phrase. One, this volume is the outline of a practical approach to understanding, and perhaps conducting, evaluations of development education programs. In this sense it is intended that the document be *worked through* – if not worked with. Two, it is being offered as *work-in-process*. We have endeavoured to present brief treatments of each evaluative stage. It is not because we believe a terse discussion is sufficient to communicate what needs to be known. On the contrary, if sufficient interest is expressed, we propose to flesh out the information required for competence in evaluation through a series of practical supplements. The point of the present handbook is to provide an impetus and a framework for identifying the evaluative concerns of the development education community. In this sense, we are tabling the document for consideration and evaluation.

TABLE OF CONTENTS

INTRODUCTION 1

STAGES IN EVALUATION 5

Stage 1. Identifying the stakeholders 6

Stage 2. Determining the stake 8
A. Solicitation 8
B. Sorting 8
i. Impact concerns 8
ii. Input concerns 10
iii. Foundation concerns 11
C. Selection 12

Stage 3. Spelling out the evaluation mandate 15

Stage 4. Picking the team 16

Stage 5. Formulating the design 17
A. Standards 17
i. Relevance 21
ii. Adequacy 21
B. Evidence 21
i. Quantitative 22
ii. Qualitative 22
C. Sources 23
i. Credibility 23
ii. Multiple corroboration 24
iii. Sample specification 24
D. Methodology 25
i. Respondents 25
ii. Subjects 30
iii. Documents 30
E. Validity 33

Stage 6. Constructing the instruments 34

Stage 7. Collecting the data 36
A. Reporting rate 36
B. Reliability 37

Stage 8. Tabulation and interpretation 38
A. Tabulation 38
B. Interpretation 41
i. Ascription 41
ii. Assessment 42

Stage 9. Reporting 43

Stage 10. Utilization of results 45

SELECTED REFERENCES 47

INTRODUCTION

Perceived deficiencies

It is our view that certain deficiencies exist which affect the well-being of evaluation within development education. The purpose of this handbook is to help improve this situation. Our perception of these deficiencies is reflected in the ways in which this handbook is structured and constituted. It is important at the outset to explain and to some extent justify the three most significant of these judgements.

- **The need to accommodate diverse characteristics of the development education movement**

Unity/diversity

To call development education a movement is to say two things. It implies common features which set it apart from other types of educational enterprises, such as physical education or art education. Simultaneously it implies a diversity of articulations or manifestations of these central features. For example, while the members of the movement share a belief in the potential of education to reform, they employ a wide range of strategies and settings. What makes it sensible to refer to these diverse endeavors by a common name is that all, in one way or another, share at least one common goal – to promote a ***global perspective*** on issues, problems and events which surround us and in which we participate. Advocates of development education hold that all of us – whether we are students in school, followers of a religious group, representatives of a political party, members of a teachers' union, owners of a small business or workers in a factory – must realize that many of the events we experience and problems we face have global implications. And further, we are often part of world-wide structures whose interactions shape the events which may otherwise appear to be merely local. However, even these shared beliefs arise from fundamentally different ideological stances. For some the underlying value position is religious, for others it is secular humanitarianism, and still others are motivated by socio-political considerations. Regardless of these different orientations, settings, or constituencies, development educators want to know whether or not the work they do is worthwhile and effective. They want, in short, to be able to evaluate their programs. Needed, therefore, is a tailored approach to evaluation sufficient to encompass a diversity of programs and a commonality of concerns.

- **The need to promote a defensible conception of evaluation**

Objectivity/subjectivity

Cynicism about the objectivity of evaluation is prevalent. This perception is

potentially misleading and warrants careful clarification. Evaluation renders an explicit ***judgement*** about the worth of a program, appraising its value or assessing its goodness against a ***standard*** of some kind. These judgements are made on the bases of ***evidence*** collected. In other words, the point of collecting evidence is to determine whether the relevant standards have been met. For example, an evaluation of car safety begins with acceptance of relevant standards or criteria – road handling ability, protection in the event of accident, tire and brake performance, and so on. Information is collected (in this case through tests) to assess the degree to which the standards are satisfied. The determination of standards, therefore, is at the heart of evaluation. Something is a standard only if it is held to be an important feature or outcome of the object being evaluated. This lack of neutrality in the determination of standards (i.e., standards are what persons consider important about the object) has fuelled a widely-held cynicism about the lack of objectivity in evaluative judgements. Perhaps we can use the example of car safety to temper this misleading view of the place of values in evaluation.

Sources of disagreement

Does the fact that two groups can disagree about the safety of a car (or the value of an educational program) mean that evaluative judgements are primarily subjective? Rather than immediately accepting a relativistic answer we should look for other explanations for this disagreement. An initial suggestion is that the groups might actually be disagreeing about facts surrounding the judgement. For instance, group A may not have access to some of group B's information. Or they may have found different results in their tests. In either case, the dispute could be resolved by sharing their information and jointly redoing their tests until they accepted the same body of evidence. If they continued to disagree (after accepting the factual evidence) then the source of the problem may be the standards for safety. Possibly the dispute rests on an interpretation of the standards, or the minimal level of performance required to satisfy the standards. For instance, group A may be satisfied with tire performance even if there was the occasional accident as long as there was little chance of injury to occupants of the car. On the other hand, group B might insist that the tires provide accident-free performance. If that is the case, how far are we really apart when A says "the car is safe" and B says "the car is not safe"? The problem here is one of standards emanating from different perceptions of the conditions for safety. Group B believes that the car must be free from damage before we can legitimately call it safe, whereas group A believes that only the occupants must be free from injury for the car to be safe. We would be better off specifying what we meant by safe (i.e., "accident and injury-free" or only "injury but not accident-free") rather than leave the impression that evaluations are primarily subjective. Perhaps it may be suggested that the subjectivity of evaluation resides in the acceptance of different standards for making judgements. For example, group A (representing XYZ Tire Company) may not include tire performance as a criterion of safety. In this case, the problem is one of vested interest. And, we would then not want to say that group A's evaluation was as 'valid' as group B's. Rather we would want to show why tire performance was a relevant standard (perhaps by pointing out the incidence of accidents attributable to faulty tires).

This example should not be interpreted to suggest that we will or should always agree in an evaluation. There is often fundamental disagreement about the appropriate standards for judging an issue. However, the point remains that it is

always appropriate to ask for justification of judgements and that some justifications are better grounded than others. Those that are better grounded can be defended by verifying the factual evidence and by validating the relevance and adequacy of the standards. If the state of evaluation is to be improved, it is important that disquieting skepticism about the impossibility of objectivity be replaced with critical interest in the justifiability of each evaluative judgement.

• The need for evaluation guidelines for non-professional evaluators

Direct participation

Professional evaluation of progams is a substantial industry, with elaborate procedures, cadres of specially-trained evaluators and a burgeoning literature. In the face of this specialization and expertise it is often felt that the non-professional has at best a minor role to play. This need not (and should not) be the case. With a basic understanding of evaluation, members of development education programs can occupy important positions in evaluations of their programs. At the least, they should assist in setting the initial course of the evaluation. Often questions about direction are answered by default by the external evaluator as the evaluation proceeds. It is not sufficiently explicit for a client merely to ask for "an evaluation". That may cover anything from a justification of ideological underpinnings of the program to the isolation of long term unanticipated consequences of a program. Neither may be of use or interest to the concerned parties. Greater understanding, and hence input, may mean that evaluation takes on a ***formative*** emphasis (namely, that a primary focus of the evaluation is to improve the program).

The extent of non-professional participation in evalution is not restricted to the role of consumer (even an enlightened consumer). Certainly, there will always be on-going need for external evaluators to carry out independent audits. However, greater understanding of the procedures will reduce reliance on external evaluators. Non-professionals can learn to perform many evaluative functions using the external evaluator more economically as a consultant. Particularly where the results are for internal purposes, non-professionals can supervise and conduct useful evaluations of their own. Even in the case of external evaluation, the ability to assess the soundness of evaluative judgements and to utilize results are enhanced if the parties understand the procedures by which the findings were reached.

STAGES IN EVALUATION

We offer, somewhat arbitrarily, ten stages in evaluation. While there is a rationale for the presented sequence there is no strict chronological order. For example, an evaluation team may be chosen prior to identification of all the stakeholders. Perhaps preliminary analysis of data reveals further questions which should be investigated. However, by proceeding through these stages, development education groups hopefully will gain a more complete understanding of evaluation. With this ground work, we may begin to repair some of the perceived deficiencies in evaluation for development education.

TABLE 1: STAGES IN EVALUATION

1. Identifying the stakeholders
2. Determining the stake
   a. Solicitation
   b. Sorting
   c. Selection
3. Spelling out the evaluation mandate
4. Picking the team
5. Formulating the design
   a. Standards
   b. Evidence
   c. Sources
   d. Methodology
   e. Validity
6. Constructing the instruments
7. Collecting the data
   a. Reporting rate
   b. Reliability
8. Tabulation and interpretation
   a. Tabulation
   b. Interpretation
9. Reporting
10. Utilization of results

STAGE 1. IDENTIFYING THE STAKEHOLDERS

In any development education program evaluation the following roles may emerge:[1]

animateur – the group leader(s), teacher(s), advisor(s), activist(s) whose role it is to put the program into use.

developer – the person(s) who have developed or are developing the program.

external decision-maker – the official(s) and agency(ies) such as school administrators, government agencies, funding bodies, educational publishers who make decisions about whether the program should be implemented, funded, promoted.

evaluator – the person(s), group(s), agency(ies) who are to carry out the evaluation.

general audience – parents, co-workers and anyone from the larger constituency who are likely to come in contact with the program, and who may care about the results of an evaluation or may indirectly influence decisions. For instance, although school administrators and teachers are responsible for accepting/rejecting a program, their decision may be influenced by parents and other community members.

internal planner – the person(s) or board of persons within the program responsible for setting the long-term course and priorities of the group.

manager – the person(s) within the program who must adminster the operations of the group on a day-to-day basis.

participant – the student(s), worker(s), adult(s), group(s) who are intended to be directly affected by the program.

All of these players potentially have a stake in the evaluation. By this we mean, persons in each of these roles may be affected to varying degrees by the evaluation. Since some persons may only have marginal interests at stake, identifying who the principal stakeholders are is an initial task in program evaluation.

Identification of principals

Sometimes identification of the principal stakeholder(s) is straightforward. For example, developers of a program may want to revise certain aspects of the learning materials and activities. As the individuals who most need the evaluation (in order to make informed decisions about revisions) they are the principal

[1] Because the institutions within which development education is carried out are so varied, the authors were faced with finding terms which would not appear to restrict the use of the handbook to any one type of institution. If we were to discuss evaluation as an activity involving teachers and students, for example, a range of unintended assumptions might have been inferred by the reader. We have chosen, therefore, to employ the most generic and hopefully least "institution-bound" terms. It should also be noted that a particular person may, as the saying goes, "wear several hats" and thereby fall into several of the roles.

stakeholders. More frequently there are a number of interests at stake. For example, at the conclusion of a particular project there may be a plethora of stakeholders. The developers of the project may want to know what worked and what didn't. Participants who have been criticized during the program might want to be reassured. External decision-makers (e.g., the community administrators or the funding agencies) and internal planners might want to know if additional programs are needed and if so, of what type. Animateurs and managers might want assessment of their contributions to guide them in future projects.

There are no easy formulas for identification of and adjudication among stakeholders. However, if there are no obvious principal stakeholders, it is wise to solicit concerns about the program from as broad a base as possible. There may emerge a core of concerns common to most of the stakeholders. To these may be added some of the particular concerns of individual stakeholders.

The point of identifying the groups and agencies who are the stakeholders is to increase the likelihood that the evaluation provides answers to significant and pressing questions. If this stage is skipped, a few individuals (including the evaluator who may have the least stake in the program) will likely determine the areas for evaluation. Money may be spent producing answers to questions that no one really cares about, or a sizeable portion of the constituency may be disgruntled with the direction of the evaluation.

STAGE 2. DETERMINING THE STAKE

It is crucial that stakeholders determine those concerns which will direct the evaluation. This task may take considerable time if the stakeholders are disparate geographically or ideologically. Without these impediments, this stage could be completed in a day-long planning session of stakeholders and invited guests. Essentially, there are three phases to this stage: *solicitation* of concerns about the program from stakeholders, *sorting* of concerns by type, and *selection* of those general concerns which will guide the evaluation.

A. Solicitation

Collection of concerns

Solicitation is the procedure whereby stakeholders are surveyed for the concerns (questions, problems, doubts) they have about the program. A useful method is to have stakeholders (individually or in groups) list the concerns which they consider important areas for investigation. Once these various questions are identified and listed, they can be condensed and, where necessary, clarified by their authors. The importance of clarity of concerns should be underscored. It is conceivable that stakeholders may describe their concerns in phrases such as "Is it a *good* program?" or "Is the program *successful*?". These are not very helpful in directing the evaluation because of their ambiguity. What does the speaker mean by *good*? Is the person asking whether experts in the field would rate it highly, or whether the aims of the program are commendable? What is meant by *successful*? Is the person asking if it is one of the more popular programs, or if the goals of the program are frequently achieved (without reflection on the merit of the goals)? In short, the solicitation stage should not be considered complete until there is a list of reasonably explicit concerns generated from the recognized stakeholders.

B. Sorting

Areas of concern

The next stage involves organizing the concerns by topic. This can be done by lumping concerns into general *areas* and then sub-classifying them by *aspect*. These areas are not air-tight distinctions; rather they are poles around which concerns can be clustered. These areas are the *impact of the program*, the *inputs of the program*, and the *foundations of the program*.

i. Impact concerns

Concerns about the impact of a program fall under the broad question "What effect is the program having?" Table 2 contains a range of sample impact concerns. In passing, it is worth noting that impact questions are most frequently asked about participants in a program, but can and should often be asked more broadly. For example, concerns 4, 8, 10, and 11 ask about the effects on the institutions and communities with which the program interacts. Notice also that impact concerns may address the program as a whole or may be directed to the effects of specific components or strategies. For example, concerns 6, 12, 13, and 14 ask about the effects of particulars of the program.

TABLE 2: SAMPLE IMPACT CONCERNS

(1) What do participants learn about development education?

(2) How much do participants remember three months after the program?

(3) Do participants change their consumer habits?

(4) What (if any) effect does the program have on the institutions from which the participants come?

(5) Does the program achieve its stated aims?

(6) Does explicit portrayal of hunger and poverty induce commitment to change or cynicism and apathy?

(7) Does the program have a 'spill-over' effect by increasing sensitivity to persons who are in some way less advantaged?

(8) Does the presence of development education programs in the schools affect university teacher training practices?

(9) Do participants donate more money to global development than non-participants?

(10) How extensively is information about the sponsoring agency dispersed by those who have taken one of the development education courses?

(11) Are there individuals or groups with hostile or negative reactions to the program?

(12) Is film more likely to increase acceptance of a 'global perspective' than guest speakers?

(13) What are the long-term effects on participants of an extended program compared with a concentrated program?

(14) Does the unit on nuclear proliferation have unhealthy psychological effects on younger participants?

Intended/unintended outcomes

Impact concerns can be sub-divided into two aspects: ***intended*** and ***unintended outcomes***. As the terms suggest, the former refers to the explicit goals of the program while the latter involves by-products of the program. It is potentially rewarding to consider the impact of programs from an open frame of mind. By doing so, people's attention is drawn to the possibility of desirable (and undesirable) side effects of a program. Procedurally, the strategy for evaluating achievement of intended outcomes may be different from uncovering possible unintended results. In the case of the former, the strategy would start with identification of the goals of the program by reading relevant documents or by canvassing appropriate persons. It would then be a matter of establishing acceptable standards of what would constitute achievement of these goals and seeing whether the evidence met or failed to meet these criteria. In the case of the unintended outcomes, it may often be more like throwing out as broad a net as

possible in an effort to snag potentially significant side effects.[2]

ii. Input concerns

Input concerns refer to the "goings-on" within the program. The focal question is "To what extent are the ingredients and operations of the program conducive to the achievement of the desired results?". Table 3 offers a number of examples of input concerns.

TABLE 3: SAMPLE INPUT CONCERNS

(15) Do the materials cover all the objectives of the program?

(16) Are participants bored by the discussions?

(17) Is implementation of materials done according to developer guidelines?

(18) Does church or union sponsorship influence the willingness of participants to take part in programs?

(19) What changes would animateurs like to see in the materials?

(20) How do animateurs make use of guest speakers?

(21) Is the current training of animateurs sufficient to ensure sound pedagogy?

(22) Is the readability level of materials appropriate for participants?

(23) What program goals are actively pursued and which are largely ignored?

(24) Do animateurs create a climate of trust so participants are willing to express their feelings about development?

(25) Do participants prefer simulation games to other forms of group activities?

(26) Do the materials contain factual errors?

(27) Are there sufficient participant-directed projects built into the program?

The common feature of input concerns is their reference to what goes on within the program as opposed to impact concerns which deal with what remains after the program is over.[3] Input concerns can be classified into two aspects: ***learning resources*** and ***transactions.***

[2]This is not always the case, often unintended results are predictable. In these cases, the principal task is to ascertain the nature and extent of the phenomenon attributable to the program.

[3]Often the distinction between impact and transaction is not obvious solely from the wording of the concern. For example, why isn't concern 16, "Are participants bored by the discussions?", an impact concern? It could be, depending, to put it crudely, whether we consider participant interest a means or an end. If one of the aims of the program is to get people eager to talk about development issues then interest/disinterest in discussions is evidence of achievement of an intended outcome. If, however, participant interest in discussions is of instrumental value (i.e., participants won't learn from the discussions unless they find them interesting), then it is a transaction concern.

Learning resources

Learning resources refer to the animateurs teaching the program (for example, concern 21); and the actual materials (for example, concerns 15, 19, 22, 26) and the methods (for example, concerns 16, 20, 25, 27) designed for animateurs and participants to teach and learn from each other.

Transactions

Transactions distinguish the actual in-use application and workings of the learning resources and strategies from their design. (For example, concerns 17, 18, 23, 24). The focus of this last aspect is the *in-use* program as opposed to the *on-paper* program.

iii. Foundation concerns

The final area of concerns cluster around the underlying values and assumptions of the program. Here, the operative question is "Are the underpinnings of the program sound?" Table 4 provides a sampling of these kinds of concerns.

TABLE 4: SAMPLE FOUNDATION CONCERNS

(28) Are the goals that we've set for ourselves justifiable?

(29) Are our pedagogical assumptions justifiable?

(30) Are our assumptions about participant abilities correct?

(31) Is anything within our goals inconsistent with a fundamental respect for persons? (Do we resort to psychological tricks?)

(32) Are the goals that we seek achievable through education or do they require more fundamental change?

(33) Are the goals we've set for ourselves representative of the things we ought to be doing?

(34) What is the nature of the animateur-participant relationship implicit in our program?

(35) Are we biased if we do not present opposing sides of the issues?

(36) What does our selection of topics suggest about what we consider most important? least important?

(37) Are our goals defensible in light of the broader goals of the development education movement?

The foundations of a program are often taken for granted. Focusing attention on foundation concerns requires a willingness to uncover and judge the basic defining features of a program. There are three aspects to foundation concerns: *program values, factual assumptions* and *basic principles.*

Program values

Program values are explicitly or implicitly reflected in goals, materials and methods. Inescapably developers make judgements about what is worth knowing

or believing, what attitudes are important, and what people should do in some area of their lives. These judgements may take the form of ***explicit*** value claims or prescriptions. Goal statements or objectives are obvious examples of this: "Participants should learn about the causes of world hunger"; "Participants will be able to detect bias in newspaper reports"; "Instructors should concentrate on group discussion and resolution of issues". Less obvious are the values ***implied*** by the judgements about what activities and topics are included in (and excluded from) the program, and how these are represented. Commonplace examples of implicit value stances are materials which reflect sexual stereotyping or bias. Foundation concerns about ***program values*** seek to indentify and clarify the matrix of judgements embedded in any program. (For example, concerns 33, 34, 36). Concerns about the remaining aspects are attempts to get at the presumptions behind the developers' decisions about program values. In other words, these aspects raise concerns about the justification of the rationale for the goals, materials and methods.

Factual assumptions

Questions about ***factual assumptions*** address such issues as empirical theories of how learning occurs (concern 29); presuppositions about characteristics of participants, such as language fluency, reading ability, socio-economic background, and interests (concern 29); and theories of how social/political/economic change occurs (concern 32).

Basic principles

Questions about ***basic principles*** address the fundamental moral tenets and the basic ideological stance presupposed by the program. Problems associated with the moral presuppositions would include doubts about whether the program treated people with dignity and respect. (For example, concerns 31, 35). Ideological considerations would appeal to views about how success should be measured (for example, in terms of the individual or the social unit) or how wealth should be distributed (for example, proprietary rights, labour, need). (See concerns 28, 37). If the program rationale is sound, it should be possible to demonstrate that the program rests on a set of principles that are acceptable and a set of factual assumptions that are defensible.

C. Selection

The point of sorting stakeholders' concerns into areas and then further into aspects is to facilitate selection of those concerns which will become the aims of the evaluation. Ideally, every concern could be investigated. Unfortunately, conducting a proper evaluation is consuming of both time and resources. Diminishing return is very much a factor in evaluation: the broader or more numerous the aims, the less evaluators may be able to adequately do the job. It is good sense generally to evaluate the subset of key concerns from the universe of expressed concerns.

Summarize

A first step in selection may be to summarize the clustered concerns into more general statements.[4] These statements should reflect, in more or less general terms, the particular areas and aspects of concerns just discussed. It should not be presumed that the stakeholders are aware of all the specific areas for which

[4] This may not apply to all concerns. In some cases a specific concern may capture a particularly important issue which should not be diffused by being lumped with other concerns.

evaluation would be useful. Rather, the concerns they express provide the direction(s) the evaluation takes. In the course of studying these concerns more carefully, evaluators are likely to uncover further concerns that had not previously been considered. These unanticipated issues may be the illuminating outcomes of the evaluation.

Priorize

Selection of the aims of the evaluation from among the potential candidates can be effected in several ways. If there is a small number of statements (less than fifteen), the consituency of stakeholders could be asked to ***rank*** them (e.g., list in order of importance each of these general statements). If there is a larger number, ***rating*** may be preferable (e.g., assign each statement to a category).

Example: rating categories

(2) Essential that it be evaluated
(1) Important that it be evaluated but not essential
(0) Not important enough to warrant evaluation at this time.

After rating, if there is not a clear consensus as to which few statements are essential, it may be desirable to have the stakeholders rank the top contenders. This list of aims in order of priority may assist the evaluators in developing the research design. Alternatively, it may be left up to the evaluation team or a committee representing the varied interests to negotiate the aims. This latter procedure may facilitate a compromise position with the selection of one or more shared aims supplemented with a set of narrow concerns. In selecting the aims, the obvious and, hopefully, the deciding criterion is the importance of the concerns to the stakeholders. There are, however, three other factors which may also have to be considered. For want of better terminology they can be called ***practical***, ***political***, and ***professional*** considerations.

Practical considerations

Practical considerations deal with the feasibility of the evaluation aims. Given the competencies of the evaluators, the time and resources available, the accessibility of the sources of information and the volume of required data, is a competent evaluation likely? If the answer is no, a narrowing or change in focus is required until it is possible to give a positive response to the question.

Political considerations

Political considerations encompass the motives of the stakeholders and the sensitivity of the potential outcomes. It is naive not to probe the reasons why each stakeholder wants an evaluation. Sometimes the real motive is to discredit another party in or to absolve some party from blame. Even without ulterior motives, results of an evaluation often are disruptive. It may be possible to predict that the evaluation, if conducted as recommended, would serve only to polarize a group or would jeopardize the continuation of what was clearly a valuable service. If any of these are suspected, then the stakeholders should reshape the purpose for which the evaluation is done, making it more justifiable on ethical and political grounds.

Professional considerations

Professional considerations refer to expert or authoritative judgements about the directions an evaluation should take. Stakeholders may want what they don't need. The mere fact that investigation of particular concerns is desired by stakeholders does not establish the value of undertaking the investigation. Often an outside expert will be able to spot misdirections which are not apparent to

persons with lesser experience or objectivity. [5]

[5]As a matter of course, it may be wise to get a second opinion, as it were, at the end of each stage of the evaluation process.

STAGE 3. SPELLING OUT THE EVALUATION MANDATE

After the aims of the evaluation have been set, there still needs to be an elaboration of other terms of reference. This ***evaluation mandate*** is the set of instructions given to the evaluator(s) to direct the evaluation.

General aims

As was just mentioned it includes statements of ***general aims*** which may or may not be priorized.

Specific concerns

The ***specific concerns*** may also be included. These would be the detailed concerns identified by the stakeholders which fell within each of the accepted aims of the evaluation. Including these specific concerns is one way to clarify to the evaluators the intentions behind the aims. Also, as will be seen, they provide a starting point in the determination of standards for evaluation.

Time frame

A ***time frame*** for the evaluation should also be suggested. This may be determined by a specific funding deadline, an annual meeting or the like. In the absence of a mandatory reporting date, a target date should still be set, even if it allows for some leeway. Any intermediate dates, such as times when programs can be observed, or periods when animateurs, participants or developers will be available may also be included.

Sources of information

Sources of information might also be suggested. This would include names and locations of persons, groups and agencies and titles of materials, documents and articles which should be consulted. These sources could be rated as essential or merely recommended.

Reporting instructions

Reporting instructions, if any, would identify the group or audience for whom the evaluation is intended. It might suggest a format and medium of presentation for the major report and could request that a shorter executive summary also be prepared.

Budget

The ***budget*** may be specified. It is likely, at this stage, that only an upper level sum indicating the maximum which stakeholders were prepared to invest would be known. At some later date, when more of the details of the design were known, the evaluators might submit a detailed projection of costs.

Contact person(s)

Finally, the mandate might identify the ***contact person(s)*** for the evaluation. This would be the person who monitors the progress of the evaluation, represents the stakeholders, and acts as a resource person for the evaluators.

STAGE 4. PICKING THE TEAM

The next stage concerns who should conduct the evaluation. The choice may be obvious – a person or group within or outside the agency may be well-qualified, well-situated and well-disposed. More likely, the duties will have to be shared. For example, a manager may be charged with developing the appropriate procedures; volunteers and animateurs may collect the data; and an outside expert may then analyze and interpret the information. There are several considerations which may assist in the selection of the best-available mix of evaluators.

Selection considerations

- *Every stage has to be done well.* If there is incompetence or carelessness in design, implementation or analysis then the results of the entire evaluation are compromised. If no one within the organization is capable of completing a particular stage well, then hire someone from outside who can. It is false economy to jeopardize the entire chain of evaluation efforts because of one weak link.

- *Impartiality is indispensable.* People have a disposition to see and hear what they want to. It is for this reason (in addition to particular expertise) that outside evaluators often are called in. This does not mean that persons who care about how the evaluation turns out cannot participate. In fact, it is important to involve persons who will eventually be utilizing the results of the evaluation. Rather, don't involve anyone who has difficulty suspending their own interests and, as a rule, try not to employ persons in tasks involving their immediate concerns. For example, the person(s) who developed the materials should not be solely responsible for analyzing them.

- *Each stage emphasizes particular evaluator characteristics.* Personalities, work habits and intellectual fortes should be considered when deciding who should do what in the evaluation. For example, the design phase and the construction of the instruments require an ability to systematically translate the evaluation aims into discrete questions. Data collection, particularly if interviews are used, requires someone who can place people at ease and listens well. The coding and verification of the data calls for plodding consistency. Interpretation of the results needs open-minded yet cautious induction.

- *All professional evaluators are not alike.* In cases where an outside evaluator is desired, be sure to select carefully. Recruit persons who come well recommended by credible referees. Look for an evaluator with direct knowledge of development education or considerable expertise in related fields. Carefully examine several previous evaluation reports to see the type of approaches used and the quality of work. And finally, pick someone who is willing to investigate the stakeholders' concerns.

STAGE 5. FORMULATING THE DESIGN

Evaluation design

The *evaluation design* is the game plan of what is required to complete the evaluation and how best to get that information. Reduced to the barest of essentials it involves answering four questions of each of the evaluation aims:

(1) What are the *standards* or criteria of worth/merit?

(2) What *evidence* is needed to establish whether or not these standards have been met?

(3) Who/what/where are the *sources* of this information?

(4) What *methodology* will best secure this information from these sources?

The interrelation of these questions is depicted in the following flowchart.

AIMS—>STANDARDS—> EVIDENCE—>SOURCES—> METHODS

Once the aims which the evaluation is to assess have been set, then standards for achievement of those aims must be agreed upon. Next, the evidence needed to establish achievement or lack of achievement will have to be specified. Following this, the sources from which this evidence will be collected should be identified. Finally, the methods by which the sources will be tapped should be formulated.

In first answering these questions, evaluators may generate lists of conceivable responses. These responses should then be refined by eliminating infeasible, inefficient or non-essential components and, as we shall later see, by checking for validity.

A. Standards

An evaluation is a judgement based on standards of worth or merit. A *standard* is a feature or attribute of the class of persons or objects being evaluated that is deemed to be a factor in determining quality, success, or value. In evaluating, say, an employee, the following are some of the standards one might use: work produced, punctuality, cooperativeness, initiative, relations with other employees and so on. In program evaluation standards by which each evaluation aim will be judged must be identified and justified. There is no point to the pronouncement that a program is a good one unless we know what standards were used and unless we agree that the standards were the appropriate ones.

Identification of standards

The task of identifying potential standards can be started by considering standards or criteria that are relevant to educational programs generally. For example, Table 5 reflects a set of defensible standards for evaluating any learning materials.

TABLE 5: CRITERIA FOR LEARNING MATERIALS EVALUATION

1.0 OBJECTIVES

1.1 *Clarity:* Are the objectives clearly stated and easy to understand? Are they kept within the instructor's view at all times?

1.2 *Completeness:* Is there a complete and explicit statement of objectives? Is there an obvious imbalance among the kinds of objectives emphasized?

1.3 *Scope:* Is the scope of the objectives too general or too narrow?

1.4 *Appropriateness:* Are the objectives appropriate to the grade level and to a range of participant interests and abilities?

1.5 *Realism:* Are the objectives achievable within the time and resource constraints?

2.0 CONTENT

2.1 *Accuracy:* Are there misstatements or omissions? Is there evidence of inaccuracies? Is there evidence of bias? (ethnic, religious, political, sex roles, multicultural, regional, occupational)

2.2 *Currency:* Does the content have currency and is it likely to remain current? Are the styles, examples, and expressions dated?

2.3 *Congruency:* Does the content match the stated intents?

2.4 *Readability:* Is the level of reading difficulty (vocabulary, style of presentation, sentence structure) appropriate for participants?

2.5 *Interest:* Does it appear to be interesting, attractive, meaningful and relevant to participants?

2.6 *Organization:* Is the content well organized? Are the ideas clearly stated and sequenced? Is it easy to understand?

2.7 *Variety:* Are a variety of resources and materials suggested or provided? Is a bibliography provided to facilitate further teacher planning?

2.8 *Significance:* Is it worthwhile for students to pursue? How can the time alloted to the program be justified?

2.9 *Depth:* Is it treated in sufficient depth and detail to be adequate?

2.10 *Redundancy:* Is the content new rather than redundant to participants?

3.0 METHODOLOGY

3.1 *Variety:* Are a variety of learning and teaching strategies suggested?

3.2 *Consistency:* Does the methodology match the stated intents?

3.3 *Flexibility:* Are alternatives suggested for different teaching and learning styles?

3.4 *Level of Questioning and Thinking:* Are questions restricted to the literal level? (Remembering or recognising of the "what, who, where, when.") Are there questions at the application level? (Applying, analyzing, synthesizing, creating, hypothesizing: why? how? what if?) Are there questions at the evaluative level? (Judging and justifying)

3.5 *Creativity:* Is creativity encouraged?

3.6 *Student Involvement:* Is participant decision-making encouraged? Does it incorporate the experiences they bring to the classroom? Are they interested?

3.7 *Individualization:* Does it allow for individual pacing rather than group pacing? Must all participants do the same thing at the same time in the same way?

3.8 *Open-Endedness:* Does it encourage a variety of participant responses rather than restricting responses? Is it divergent rather than convergent?

3.9 *Evaluation:* Do evaluation strategies accommodate participant differences? Are they matched with intents rather than content?

4.0 ORGANIZATION

4.1 *Linkages:* Are there clear and appropriate linkages among the learning activities? Why are the activities sequenced in the way they are? Cummulatively, do the activities lead somewhere?

4.2 *Development:* Does one learning activity lead to the next in a logical and progressive development towards realization of the program's goals and understanding of its content? What learning and pedagogical principles are evident in the organization of the materials?

5.0 CONSISTENCY

5.1 *Internal Consistency:* Is there consistency between and among the objectives and teaching/learning activities?

5.2 *External Consistency:* To what degree are the objectives congruent with the official curriculum guidelines? Are the strategies, materials, and content supported by current informed opinion regarding: what constitutes good teaching practice; child development and learning theory; valid knowledge; etc.

5.3 *Social Significance:* How significant is the content selected for participants living in a geographically diverse society characterized by various interest, cultural, occupational, and religious groups? Why?

In addition to general educational factors, other standards specific to development education should be considered and listed. Any feature that affects the achievement of the particular goals of the program is likely a standard for input concerns. All of the goals of a program are prime criteria for impact concerns. In the case of foundation concerns, standards can be identified from the goals and rationales of the broader development education movement. For example, the goals of a particular program can be evaluated in light of their consistency with the

goals of the development education movement and the rationales for these goals. A sampling of goals and rationales for development education is reflected in Tables 6A and 6B.

TABLE 6A: FOUNDATION STANDARDS
REPRESENTATIVE GOALS OF DEVELOPMENT EDUCATION

1. Knowledge and understanding of:
   a. substantive development issues
      1) causes
      2) problems
      3) solutions
   b. abstract development issues
      1) concepts
      2) principles
      3) theories

2. Skills in:
   a. information gathering
   b. critical assessment
   c. leadership/participation

3. Fraternity/empathy/concern

4. Commitment/action

TABLE 6B: FOUNDATION STANDARDS
TYPICAL RATIONALES FOR DEVELOPMENT EDUCATION

1. Ethical reasons
   a. halt/reverse gross inequity
   b. promote justice/fairness
   c. respect human rights/dignity
   d. less ego/ethnocentric world view

2. Prudential reasons
   a. shared problems
   b. solutions require cooperation
   c. national self-interest
   d. coping with world change
   e. more enlightened/future-looking decisions

Justification of standards

Once this enumeration of standards is completed, an attempt at justification should be made. This involves determining the *individual relevance* and the *collective adequacy* of the standards.

i. **Relevance**

A standard is relevant if it is a factor in meeting the function or aim for which the object is intended. For example, in determining whether a pen is a good one, legibility, weight, and shape are all relevant standards because they bear on the capacity of a person to use the pen. On the other hand, being waterproof is not normally a relevant standard (unless the pen is intended for use in inclement weather). Comparably, in determining the relevance of standards for educational materials we must ask if they affect the achievement of the aim for which the materials were intended. Obviously, materials are designed to help achieve some or all of the objectives of the program to which they belong. Any factor which affects these outcomes is a relevant standard. Cost of materials, for example, will be relevant only if the materials would be used more or less often at different prices. In some situations the relevance is contingent on the nature of the audience. For example, appearance, colour, abundance of pictures and the like are frequently deemed to be relevant features of materials for younger students and not so for older ones. The presumption is that older students can attend more directly to the text while younger ones need frequent inducements to recapture their attention.

ii. **Adequacy**

Assesing for collective adequacy is an attempt to establish whether the set of standards provides sufficient grounds for making a judgement. This does not require that every relevant standard be included. Rather, the concern is that none of the more significant standards have been overlooked. The implication that not all standards are of equal value should not be missed. Standards admit of degrees of importance depending upon how central they are to the achievement of the program functions or goals. In justifying the standards for evaluation, it is important to ensure that at least all the important ones, and only relevant ones, have been included in the assessment.

B. Evidence

After determination of standards for evaluation of each of the aims, the type of information which will allow us to verify whether the standards were met must be identified. For example, if participant interest is considered a relevant standard, what will count as evidence that it has been achieved? Obviously, the way participants reacted to the materials is an indication. This may refer to oral statements made or by actions such as participation in activities, completion of assignments, or volunteering for additional assignments. Evidence of participant interest could also be offered by demonstrating that the topics and treatment employed by the materials conformed to patterns that have proven popular with similar audiences. There are often a number of pieces of evidence that will indicate that the standard has been achieved.

Types of evidence

The classic distinction made between types of evidence is that of *quantitative* and

qualitative data. As the names suggest, the force or weight of the information derives from different features.

i. Quantitative

With quantitative data, the value of the evidence is determined by the number of persons/units represented in each category. For instance, if we wanted to assess the effects of development education activities in a particular community, the following quantitative evidence might be sought:

- number of volunteer hours plotted over time or compared with results from similar communities without development education activities;
- per capita donations of money and goods to development projects compared with results from similar communities without development education activities;
- growth in development group membership;
- students' scores on a test of general knowledge of development issues compared with students from similar communities without development education activities.

ii. Qualitative

With qualitative data, the value of the evidence is determined by the insight and perspective of the person(s) providing the information. For instance we would look for qualitative evidence of the effects of development education activities on a particular community from persons in strategic positions. This might include:

- talking about the problems in recruiting volunteers with development personnel in several communities;
- discussing development issues with students and citizens in various communities;
- participating in and reporting on the delivery of programs or the conduct of public meetings on development issues in various communities.

Complementary methodologies

As these examples imply, the difference is not one of competing methodologies. Rather, it makes sense to view quantitative and qualitative data as complementary. Each has its own strengths which enhance our ability to draw defensible conclusions. Good quantitative evidence allows the evaluator to generalize with confidence about the population and provides data in a form that can be averaged, correlated, and otherwise statistically analyzed in ways that help to digest the information. Good qualitative evidence captures nuances, contexts and the flavour of a particular situation and is open-ended – allowing one to probe and discover in ways that pre-arranged, standard questions do not. It is recommended that quantitative and qualitative evidence be used in tandem. For example, pen and paper questionnaires and checklists can be used to quantify changes in attitudes resulting from a program, increases in knowledge of participants, and perceptions of a program. Follow-up interviews can be held with some of the participants from each program. These partly-structured meetings might be used to clarify and explain differences in responses between programs and to gain greater understanding of the reasons/feelings behind the written responses. The addition

of anecdotal comments, detailed explanations and background information will likely enrich the evaluators' ability to interpret what the 'figures' are saying. In fact, some conclusions that might have tentatively been inferred from analysis of the quantitative data alone may be shown by the qualitative data to be misleading. In short, using both types of evidence increases the likelihood of capturing a more complete and accurate picture.

C. Sources

Types of sources

Locating the best sources for securing the evidence which will show that the relevant standards are met is the next phase of design formulation. There are essentially three types of sources of evidence:

- respondents (personal reports of participants, animateurs, developers, and so on)
- subjects (behaviour of persons in situations such as classrooms and group meetings)
- documents (learning materials, minutes of meetings, articles, reference books, and so on)

Concern for accuracy

Beyond the obvious question of the accessibility of these sources, the overriding concern is accuracy. Will the information obtained from the particular source reflect what really happened and will it tell us what we want to know? Some of the factors which influence accuracy are the *credibility* of the reports, use of multiple sources to *corroborate* evidence, and method of selection of the *sample population.*

i. Credibility

Types of criteria

Assessing credibility involves rating sources on the following types of criteria:

- *Was the person in the best position to report on the event?* There are a number of evidential rules of thumb implied by this criterion:
  - other things being equal, first-hand reports are more accurate than second or third-hand reports;
  - other things being equal, in reporting on internal states (thoughts, emotions, feelings) the person who felt/experienced them is a more trustworthy source than a mere witness;
  - other things being equal, a trained observer is a better source than a participant who may have been too caught up in the activity to see what else was happening;
  - other things being equal, the sooner the report is delivered after the event the better.
- *Is there evidence that the person has a predisposition which would bias the report?* This criterion goes beyond the rejection of obviously prejudiced sources. Less apparent for example, is the questionable use of young children as sources of evidence. The desire of many children to please authority figures means that often they answer what they think will most please the

interviewer and not what was actually seen, done or felt.

- *Does the person have a good reputation, particularly for truth-telling and accuracy?*
- *In the case of authoritative sources, has the person studied the particular issue in some depth or had considerable experience in the matter?*

ii. Multiple corroboration

Types of corroboration

The use of multiple sources of information is a way to verify the accuracy of information or, at least, to flag the likelihood of misinformation. There are two versions of this test: *internal consistency* with the same source or *external triangulation* with different sources.

1. Internal consistency

Corroboration with the same source may involve repeating the same or similar question to a person at some later date. Alternatively it may involve asking contemporaneous questions in various forms which essentially solicit the same information.[6]

2. External triangulation

Corroboration with external sources involves using different sources to confirm various others' reports. For example, if we wanted to determine whether the reading level of learning materials was appropriate for the participants, the following different sources may be explored: statements by animateurs and participants and comparison of participant scores on standardized reading tests with the assessed readability level of materials on standardized analysis tests. External corroboration (and internal consistency) may include multiple use of either quantitative or qualitative evidence or (perhaps ideally) could include a mix of both types.

For every piece of evidence required, one or more types of sources may be tapped. In the early stages, it may be advisable to identify as many credible sources as possible. As the design is refined, sources may be eliminated. For example, in finding evidence to verify participant interest, initially participants, animateurs and an observer may be identified as likely sources. Later it might be decided that the observer's time would more profitably be spent collecting data that was not otherwise available, leaving the animateurs and the participants as the sole sources.

iii. Sample specification

After the general sources of information have been specified, the actual persons, materials and situations to be sampled must also be specified. In large scale surveys, *population sampling* is done to select those which are representative of all potential sources. In studies involving a small number of sources the presumption is that evaluators will include the entire population. However, if sampling is desired because, for example, there is enough time to observe only a few of the sources, then the selection should be made on the basis of the following considerations. First, the larger the number, the greater the likelihood that any

[6]Use of internal consistency is usually reserved for occasions where a particular person or group is the only accessible source and there is a suspicion that the respondent may be hesitant to come right out with the answer.

significant differences will be covered. Second, when making the selection, choose by one of the following methods:

Methods of sampling

- *random*: let chance decide which sources are included by picking numbers or sorting cards; or by
- *cross-section*: identify the features that are likely to be factors in the success of the program (e.g., type of participants, animateur teaching styles, learning materials used) and select so that there is a program to reflect each of the important variables; or by
- *typical cases*: identify sources that within themselves seem to encompass a broad, representative spectrum of characteristics.

D. Methodology

The final phase of evaluation design is specification of the method(s) to be used to secure the required evidence from the identified sources. As was mentioned in the previous section, there are basically three sources of evidence: ***respondents*** to the study, ***subjects*** in the study, and ***documents***. In simplest of terms, the methodology for each is as follows: respondents are questioned, subjects are observed, and documents are examined. Information is gathered from a respondent by asking questions and verbally collecting their deliberate responses. Information is gathered from a subject by watching what the person does. The point of treating sources as subjects is to collect information about context-related behaviours.[7] Information is gathered from documents by examining them. The point of examination is to expose the constituent parts of the document.

i. Respondents

Some of the more obvious ways to collect information from respondents are itemized in the Table 7.

Open-ended format

As indicated in Table 7, there are two formats for data collection methods for respondents. An ***open-ended format*** implies that the respondent answers in his/her own words. For example, in an interview or on a questionnaire, a respondent may be asked simply, "Were you satisfied with the help you received from your animateur?", or "Describe the help you received from your animateur."

Closed format

The ***closed format*** specifies that the respondent answers by selecting one or more of a prearranged set of responses. The distinction between open-ended and closed formats is perhaps best viewed as poles on a continuum. Open-ended questions can be explicitly directive to the point that they operate much like closed questions. For example, the open-ended question, "How often did you miss meetings?" narrowly proscribes the range of response. It is, strictly speaking, still open-ended because respondents could answer in absolute terms (I missed three meetings), in relative terms (I missed 10% of the time), or in comparative terms (I missed as many as the average person).

[7] An animateur would be both a subject and a respondent if an evaluator, after observing the person in the classroom, asked questions of the person about the lesson.

Example: closed format

Were you satisfied with the help you received from your animateur?
_____(1) Very satisfied
_____(2) Quite satisfied
_____(3) Somewhat dissatisfied
_____(4) Very dissatisfied

TABLE 7: DATA COLLECTION METHODS FOR RESPONDENTS

Medium of communication	*Format*	
	open-ended	*Closed*
1. ORAL		
A) Interview	X	X
- the gatherer invariably asks the views of the respondent		
Evaluator ——> Respondent		
B) Discussion	X	
- the gatherer and respondent share views		
Evaluator <——> Respondent		
2. WRITTEN		
A) Questionnaire	X	X
- a set of specific questions whose answers are coded as to type of response		
B) Test	X	X
- a set of specific questions whose answers are scored for accuracy/quality of response		
C) Exposition	X	
- a longer analytical, evaluative or descriptive account		
D) Narrative	X	
- a longer experiential account		

Modified closed format

Often questionnaires include *modified closed format* questions which allow for some flexibility in responses.

Example: modified closed format

What main reason did you have for joining the training program?
(check only one)
_____(1) wanted a steady job
_____(2) needed the money
_____(3) was convinced to join
_____(4) wanted to teach
_____(5) wanted to have a leadership role
_____(6) had nothing better to do
_____(7) other (specify)_____

Each format has positive and negative features. Some advantages of the closed (or modified closed) format over the open-ended format are:

Advantages

- ease in administering the instrument
- keeps respondent on topic
- ease in tabulating results
- more uniform responses facilitate measurement and comparison.

Some of the disadvantages of closed format are:

Disadvantages

- may fail to reveal much depth/insight
- may not be subtle enough to discriminate among respondents
- may force a respondent to take an alternative that does not accurately reflect opinion
- may be more difficult to construct the instrument.

When constructing closed format questions, it is imperative that the options are exhaustive of all alternatives and mutually exclusive. The reward of going to this trouble at the instrument construction stage is that tabulation of the results is greatly facilitated. In fact, as the number of respondents to the evaluation increases, the demands of collecting, tabulating and comparing large numbers of responses invariably results in heavy reliance on closed format questions. Generally the attempt is made to minimize the trade-offs by balancing the approaches. Some strategies include:

Balanced approach

- holding open-ended interviews/discussions with all of the *key* respondents
- using modified closed format questions where feasible
- interviewing a sub-sample of respondents to a written questionnaire for a more complete understanding of the larger body of data
- using of a variety of closed format measures to improve the depth of responses (e.g., following up a true/false question with a request for ranking of potential

options).

There are a variety of suitable closed format measures. The most common forms are listed in the Table 8.

TABLE 8: CLOSED FORMAT MEASURES

1. DISCRETE CHOICES

A) Multiple choice

Q. When did you start the program?
____(1) at the very beginning
____(2) sometime later in 1978
____(3) sometime in 1979
____(4) sometime in 1980
____(5) sometime in 1981 or later

B) Yes/no

Q. Are you 25 years old or younger?
____(1) Yes
____(2) No

C) True/false

Q. The per capita GNP of Canada is greater than that of the U.S.
____(1) True
____(2) False
If the question is such that you want to discourage guessing, include a third option:
____(3) Not sure

2. SCALES

A) Rating

Q. In teaching this unit, I found the animateur's guidebook:

Useful						Useless
1	2	3	4	5	6	7

Clear						Vague
1	2	3	4	5	6	7

B) Rank-order

Q. In order of importance, with

1 as the MOST IMPORTANT
5 as the LEAST IMPORTANT

rank the following development educational goals:

____Raise general public awareness
____Increase number of local volunteers
____Increase amount of dollars donated
____Alter consumer habits
____Increase demand for political change

C) Equal appearing intervals

Q. Suppose you overhear a conversation among foreign students attending a local Canadian university. They are critical of the federal government for its support of several Canadian corporations with operations in Chile. What is your reaction likely to be?

____1) I cannot predict without knowing more about the situation.
____2) I would tell them if they didn't like what we do in Canada, they should get back home.
____3) I might not say anything, but would think they were ungrateful.
____4) I might not say anything, but would wonder why they were upset.
____5) I would reassure them that not all Canadians are supportive of the operations of international corporations.

(Note: This form of question covers the spectrum of reactions a respondent might have and attempts to provide a statement which reflects each significant position on that spectrum.)

D) Q-Sort

Q-Sort scales measure attitudes, interests and values by sorting cards. For example, one way to measure racial bias among children would be to show them pictures of activities and ask them to identify activities they would like to be doing. If two sets of pictures existed, one which depicted children of the dominant racial group alone and another set which depicted mixed or non-dominant racial children, then comparison about the relative popularity of the activities could be made.

ii. Subjects

When the sources of information are subjects, the intention is to collect data about behaviour in a contextual situation – people *in-action*, as it were. There are three typical ways to do so: *observation, experimentation,* and *participant-observation.*

Observation

Observation, as the name implies, consists of watching while a individual or group goes about its ordinary activities. Often an observer will use observation checklists or anecdotal notes to record the actions or events deemed important. Observation is often used to study transaction concerns.

Experimentation

Experimentation implies the manipulation of events so as to discover what people will do in particular situations. Often the results are compared with a group that did not receive the same treatment. For example, an evaluation experiment might have development education participants undertake a simulation about world development. Their attitudes or knowledge might then be compared to persons who had not participated in the game.

Participant-observation

Participant-observation, as the name implies, involves the evaluator in the activities of the group. The point in doing so is to help the evaluator see what the participant sees. An evaluator taking the development education program along with other participants is one example of this method.

iii. Documents

When the sources of information are documents, the point of the methodology is to identify the constituent parts and underlying themes. There are several ways of collecting information of this nature: *enumeration, review,* and *analysis.*

Enumeration

Enumeration involves counting the frequency of certain characteristics. For example, in assessing the impact of a program one could examine documents looking for the number of members in development groups, the amount donated, or the number of volunteer-hours. Counting the number and types of activities prescribed in the learning materials is a good way to unearth the *de facto* emphasis of the program.

Review

Description or *review* is self-explanatory and includes the following sorts of data collection: summary of the literature or archival records on a topic, encapsulization of participant written work, notes on the contents of materials and so on.

Analysis

Analysis is a more probing form of description and often includes evaluative judgements. The objects of analysis are digested, as it were, to expose their more covert aspects. Typically this mode of data collection is used with input concerns (learning resources and methods). Table 9 contains an extract from an analysis scheme for use with development education learning materials.

Now that our discussion of major components of an evaluation design is complete we can piece together a rough design. Table 10 presents a sample of the types of standards, evidence, sources and methods that might be considered if evaluating

the soundness of materials for use in public high schools.

TABLE 9: EXTRACT OF A MATERIALS ANALYSIS SCHEME

1. What is the number/percentage of the units/parts of the materials that focus on:

	No./%
A) Presentation of empirical facts or data (e.g., historical, geographic, descriptive)?	____
B) Discussion of issues, presentation of arguments, justification of values?	____

2. With what complexity are issues/problems presented?

A) ____ Detailed analysis with many of the complexities addressed
B) ____ Some sophistication but obviously simplified
C) ____ Compexities not addressed

3. What aspect of issues/problems is stressed?

	Major Emphasis	Secondary Emphasis	Little or no mention
A) Causes/origins	____	____	____
B) Manifestations/ramifications	____	____	____
C) Remedies/programs	____	____	____

4. Are differing positions on the issues mentioned? A) ____ Yes B) ____ No

If yes, how are the opposing positions treated?
i) ____ Presented in as favourable a light as possible with rationale
ii) ____ Views expressed with some supporting reasons
iii) ____ Dismissed or hardly mentioned

5. Is there evidence by omission or commission of bias/stereotyping?

	Yes	No
A) Class	____	____
B) Race/ethnic	____	____
C) Religious	____	____
D) Sexual	____	____
E) Other (Specify)_____	____	____

If yes, explain _____
_____

TABLE 10: PARTIAL DRAFT OF EVALUATION DESIGN

AIM: EVALUATE THE SOUNDNESS OF THE MATERIALS FOR USE IN PUBLIC HIGH SCHOOLS

STANDARDS	EVIDENCE	SOURCES	METHODOLOGY
1) INTERNAL CONSISTENCY	COMPATABILITY AMONG GOALS, METHODS, AND RESOURCES	MATERIALS	ANALYSIS
2) CLARITY OF LESSON PLANS	EXECUTION OF LESSONS	TRANSACTIONS	OBSERVATION OF LESSON(S)
	EASE OF UNDERSTANDING	ANIMATEURS	INTERVIEW/QUESTIONNAIRE
3) SUITABLE FOR STUDENTS	READABILITY OF MATERIALS	MATERIALS	ANALYSIS (READING LEVEL DIFFICULTY)
	READING ABILITY	PARTICIPANTS	OBSERVATION (READING ASSESSMENT)
4) ACCURACY OF INFORMATION	ERROR	MATERIALS	ANALYSIS
		ANIMATEURS	QUESTIONNAIRE
		PARTICIPANTS	QUESTIONNAIRE
	OMMISSION	'EXPERTS'	INTERVIEW
		MATERIALS	ANALYSIS OF COVERAGE
5) PARTICIPATION	ACTIVITY OF PARTICIPANTS	ANIMATEURS	INTERVIEW/QUESTIONNAIRE
		PARTICIPANTS	QUESTIONNAIRE
		TRANSACTIONS	OBSERVATION OF LESSON(S)
6) VARIETY OF ACTIVITIES	FREQUENCY AND TYPE OF PARTICIPANT 'ACTIONS'	ANIMATEURS	INTERVIEW/QUESTIONNAIRE
		PARTICIPANTS	QUESTIONNAIRE
		TRANSACTIONS	OBSERVATION
		MATERIALS	ENUMERATION

E. Validity

The concluding, and likely overriding, concern in the refinement of the design is *validity*. For our purposes the relevant measure of validity is what is called ***face validity***. As the phrase suggests, the assessment involves asking if, ***on the face of it***, the set of proposed methods will do the job they are intended to do. This is an assessment of the extent to which the type and weighting of evidence-gathering methods adequately cover the standards which must be addressed. For example, will data collection methods (A, B, C) from sources (D, E, F) collecting evidence (G, H, I) about standards (J, K, L) allow us to draw defensible conclusions about the value of the project? The four links in the verification for validity of the design are reflected in the following diagram:

METHODS —>SOURCES —>EVIDENCE —>STANDARDS —>AIMS

In testing for validity, we must confirm that each of the links completes its function. One, are the methods chosen likely to elicit accurate information or are they rather crude measures, inappropriate for the designated sources? Two, are the sources of information varied and credible? Three, does the quality and quantity of evidence collected allow us to defensibly conclude about the degree to which the standards have been met? And finally, are the standards individually relevant and collectively adequate to judge the program?

STAGE 6. CONSTRUCTING THE INSTRUMENTS

Item construction

Once the evaluation design has been refined, the next task is to construct the data-gathering measures or *instruments*. The pieces of evidence required should be listed. A search for questions, checklists, analysis schemes and the like which may be relevant should be carried out. If appropriate models are available, start with them, as the task of building instruments is difficult. Either way, in the selection or the formulation of question items, the following suggestions are pertinent:

Suggested don'ts

- avoid loaded or emotionally charged words (e.g., Do you agree with federal support of *imperialistic* international corporations?);
- avoid superficial, pat questions, that may encourage stereotypical, uniform responses (e.g., Do you *like* attending development education courses?);
- avoid double-barrelled questions, as respondents may not know which part to respond to (e.g., Do you *sign petitions* and *demonstrate* against protective tariffs?);
- avoid questions which presume knowledge, experiences or past practices that respondents may not have (e.g., After the course, did you *continue* to buy boycotted products?);
- avoid esoteric or technical words (e.g., Are the materials helpful in the *conscientization* of participants?);
- avoid negatively worded questions (e.g., Are you *not* now boycotting South African products?).

Suggested do's

- use clearly worded, precise and unambiguous language;
- use short, uncomplicated questions;
- ask only what is absolutely necessary;

Instrument assemblage

The next stage is to assemble the pool of question items for use by the interviewer, observer, or respondent. Organize the instrument so that it is easy to follow. Begin with clear directions about the purpose of the study, what is expected of the participants, and how to complete the instrument. Start with relatively simple, directive questions and build up to the more complex, open questions requiring thought or soliciting sensitive information. Poorly reproduced, difficult to read, carelessly set out questionnaires are not likely to be effective. And, most importantly, make sure the instrument is not too long. The longer it is the less likely respondents will complete it, or if they do, the less carefully the questions will be considered.

Field testing

Before an instrument is used to collect evidence, it should be *field tested* with people similar to those who eventually will be asked to work with it. Failing that, have several colleagues complete or use the instrument. In addition to completing the questions, trial respondents could be asked direct questions about validity: "Is the wording of the questions clear?" "What kinds of information are the questions

likely to elicit?" "What do you see as the intention behind each question?".[1] On the basis of these trials, refine the instrument. Questions which don't discriminate (i.e., questions which are uniformly answered by almost all respondents) should either be sharpened or deleted. Generally, questions which are misunderstood, left blank, glibly answered, or are otherwise deficient should be revised. An added reason for field testing is to assist with the generation of acceptable answers to closed format questions. An early draft of an instrument can be left in open-ended form; from the responses received through field testing, an exhaustive, mutually exclusive set of options can be produced.

[1] At this stage of the evaluation, validity is extremely important. Do the questions actually measure what they are intended to measure? Expert advice about the validity of instruments may be desirable.

STAGE 7. COLLECTING THE DATA

Representativeness of results

Once revised the instruments are ready for data collection. One of the key considerations at this stage is concern for *representativeness* of results. Representativeness refers to the likelihood that the information provided by the respondents/subjects in the study accurately reflects the characteristics and opinions of the total population source. The problem of representativeness of results is not eliminated even if the entire population of respondents is used. There are two major reasons why information obtained from repondents and subjects may not reflect the true picture.

A. Reporting rate

Not all persons approached to participate in an evaluation will do so. The percentage of persons who take part from those solicited is called the *reporting rate*. In other words, if fifty persons are asked to complete a questionnaire and only twenty respond, the reporting rate is 40 percent. Reporting rate is an important consideration in assessing the representativeness of results. For this reason many evaluators insist on a high (70 to 80 percent) rate of return. The concern over a low reporting rate stems from the possiblity of what is called a *self-sampling* bias. The group of people who choose to respond to the survey may be different in significant characteristics than the non-reporting group. For instance, participants who enjoy the program may cooperate more than those who think less well of the program. The reverse is also possible.

Reduce impediments

To combat a low reporting rate, one needs to accentuate the reasons for participating and minimize the impediments to participation. Comments made earlier about the clarity and length of instruments are clearly relevant. As much as possible, make participation at the convenience of the volunteer. For example, hold interviews during off-hours; provide pre-stamped, pre-addressed return envelopes; and set aside time during regular meetings to complete questionnaires. Explain the purpose of the project and the importance of the specific information requested, and allay any fears volunteers might have (such as confidentiality of individual results,[9] self-incrimination, and psychological stress) by means of letter, presentation or informal contact. Have a person trusted by the respondents make the initial appeal.

Follow-up

Another means to increase the rate of return is through follow-up procedures. Non-respondents may have moved, lost the initial request, or otherwise need to be reminded. It may be wise to follow up with personal or direct modes of communication (e.g., telephone). In the event that the reporting rate is unacceptably low, securing responses to the essential questions may be done through a shortened version of the instrument, or informally over the telephone or in conversation.

[9] Participants' identity may need to be kept anonymous. Omit personal references on all answer sheets and reports; if some identification is required, use a number code rather than the respondent's name.

B. Reliability

A second consideration affecting representativeness of responses is that of the *reliability* of the information received from the respondents and observers. Reliability deals with the consistency of the results. Would respondents answer in similar ways if they were asked the questions at different times or under different circumstances? For example, were the respondents forced to complete the questions quickly or were they, in some other way, under pressure at the time? Were particular groups in an uncharacteristic frame of mind or treated significantly differently by the interviewer? For these reasons it is important that instructions and conditions surrounding administration of questionnaires and interviews be virtually identical in all cases. Reliability is also a factor when the sources of information are subjects or objects. In these cases, the principal concern is with the evaluators. With observation checklists or materials analysis schemes, are all the evaluators using similar and consistent criteria and methods? One means of checking for this reliability *(inter-rater reliability)* is to have all evaluators rate identical subjects/objects and compare results. To maximize the likelihood of respondent and rater reliability, provide written guidelines for the use of each instrument in a relatively uniform manner. The overriding concern in this stage is to ensure that information is collected under similar conditions and that it is an honest and careful representation of the entire population.

STAGE 8. TABULATION AND INTERPRETATION

As received, the information from respondents, observers and interviewers is in raw form. This *raw data* must be *tabulated* and *interpreted* before it is ready to be reported. The point in tabulating is to summarize, organize, and display the information in ways which facilitate interpretation. Interpreting means drawing defensible conclusions and inferences from the data. [10]

A. Tabulation

Of quantitative data

Two forms of tabulation of quantitative data are *frequency distribution* and *cross-tabulation.*

1. Frequency distribution

A frequency distribution reports how often a particular option or score occurs. As a hypothetical example, Table 11 shows how many hours per week on average participants spend on development work and what percentage of persons choose each of the options.

TABLE 11: FREQUENCY DISTRIBUTION

Q. Since participating in the course what is the *average* amount of time you spend per week doing development work?

	No.	%
Option 1 – no time	33	55.0%
Option 2 – up to 3 hours per week	17	28.3%
Option 3 – 3 hours or more per week	7	11.7%
Option 4 – No response	3	05.0%
TOTAL	60	100%

Table 11 shows that over one-half the participants (55.0%) spend no time doing development work and over one-quarter (28.3%) spend up to three hours. Let's further suppose that this program offered four separate courses ("resource sharing," "food production," "international trade" and "literacy") and that the enrollment for each class was 15 participants (N=60). If we wanted to compare how much time participants in each course spend on development work we would need a *cross tabulation.*

2. Cross-tabulation

A cross-tabulation is the frequency of the occurence of one set of options matched with the frequency of a second set of options. The point of a cross-tabulation is to identify relationships between variables. Following with our hypothetical

[10]These discussions are extremely simplified. They do not address statistical methods of manipulation and analysis of data. Rules of inference and induction are not treated. The modest point of this section is to provide a few fundamental instructions on how to summarize and read raw data.

example, Table 12 indicates that most participants in the course on international trade spend no time doing development work. The table also indicates that the vast majority of people doing development work participated in either the resource sharing or food production courses.[11]

TABLE 12: CROSS-TABULATION

	Course Participation By Development Work				
TIME	no time	0-3 hrs	3 or more	no response	
COURSE					TOTAL
Resources	6	6	2	1	15
Food	5	6	4	0	15
Trade	13	2	0	0	15
Literacy	9	3	1	2	15
TOTAL	33	17	7	3	60

If desired, data from open-ended questions may be coded. One method for translating unsorted data into categories is to look for emergent themes and patterns in the responses. There is no prescribed way to identify these trends, although it helps to read all responses (perhaps several times) before trying to identify appropriate categories. Whether the data are open-ended and being coded for the first time or are closed-ended and being recoded, two further considerations are worth mentioning:

- demarcation of descriptive categories influences interpretation, and
- descriptive categories (at least initially) are preferred over evaluative categories.

[11]It is important to be clear that all that can safely be read from this table is that persons who participated in some courses spend more time doing development work than persons who participated in others. It would be a matter of interpretation (not description), and a very shaky inference at that, to conclude that the courses contributed to this result. Participants in the "resource sharing" and "food production" courses may already have been doing development work (perhaps even more than they are currently doing). Perhaps the correlation has nothing to do with the courses *per se* but depends on the animateur. For example, if the animateur(s) had switched courses would the results be reversed? As it now stands, the correlation is an interesting one. Further investigation of the participants, the animateurs and the courses would be necessary (and desirable) before inferences should be drawn.

A *descriptive* category merely summarizes data; an *evaluative* category draws inferences or makes judgements about the data. To explain what is intended by the first these statement consider a body of raw data which consists of a list of results from classroom tests on participant knowledge of a development issue. If the results were collapsed into two descriptive categories, "80% or above" and "below 80%", this would create a different impression than had they been collapsed into two other descriptive categories: "50% or above" and "below 50". While each coding system summarizes the data, a less prescriptive alternative, which would summarize the data yet not create as dichotomous an impression, would be to distribute intervals more evenly over the range of scores.

Example: sample descriptive categories

(5) 90% or above
(4) 75 – 89%
(3) 60 – 74%
(2) 50 – 59%
(1) below 50%
(0) no results

As a general rule, particularly with the initial tabulation of results, use a more specific set of coding intervals. One can always collapse the categories without resorting to the raw data, whereas the reverse is not true. There is a trade–off in setting coding categories. Categories must collapse data sufficiently to be manageable; yet, not be so condensed as to diminish or prejudice interpretation.

A second consideration in coding is the type of category. Compare the following set of evaluative labels with the previously recommended set of descriptive labels:

Example: sample evaluative categories

(5) excellent command of key issues
(4) good command of key issues
(3) some understanding of key issues
(2) important gaps/flaws in understanding
(1) gross misunderstanding of key issues
(0) no results available

This latter set of coding categories (if applied to a list of raw scores) has interpreted the results whereas the former set merely summarized the scores. The evaluator may eventually want to make this interpretation. However, the translation of a raw score to an evaluative label requires justification and this will be discussed in the next section.

Of qualitative data

Strictly speaking, qualitative data are not tabulated (put into tables). If they are reduced to numbers – they lose the characteristics from which they derive their name – they become quantitative data. However it is sometimes desirable to sort qualitative information into clusters. This organization can be carried out in any combination of the following:

- *valence* – negative (critical) comments as opposed to positive (supportive)

comments;

- *topic* – comments on learning materials, separated from comments on activities, on animateurs, on scheduling and so on;
- *source* – comments distinguished by sub-groups;
- *substance* – comments arranged by the recommendations/suggestions they offer.

B. Interpretation

There are two considerations that are paramount in interpretation of data for purposes of program evaluation. One, the evaluators should show that the results are ascribable to the program (and not to factors independent of the program). Two, the evaluator should attempt (where defensible) to *assess* the significance, worth, or merit of these *ascriptions*. The task of assessment is at the heart of evaluation. However, the evaluation makes no sense unless it is an evaluation *of the program* and that requires undertaking the former task.

i. Ascription

To ascribe an occurrence to a program is to say that it is reasonable to conclude that the occurrence is associated with the program. It does not require showing that the program *caused* the result.[12] It only requires that the evaluator defend the reasonableness of assigning some influence on the result to the program. Sometimes this connection is directly established in the data. For example, respondents will often state that they learnt such and such *from* the program. Unless there is reason to suspect the claim, this type of evidence is sufficient to defend the ascription. Often, however, ascriptions must be *inferred* from the data. We saw this in the discussion of the hypothetical correlation between amount of time spent doing development work and the course taken.[13] The fact that participants in some courses donate more time to volunteer work does not mean that those courses influence this occurrence. To defend that ascription requires two types of evidence. One, indication that change has taken place is required. For example, the amount of time volunteered may not have altered. Some comparison of amount spent on development work before and after taking the course is required. Generally, measure of changes is achieved by tests or attitude scales taken prior to and after the program (*pre-* and *post-tests*) or by a questionnaire administered after the program which asks about prior activities. Two, some attempt may be made to locate the locus of change within the course. For example, is the subject matter a contributing factor? (Is the topic "resource sharing" more compelling than "international trade?"). Is it the animateur? (Would the results be altered if assignment of animateurs were reversed?). Is it methodology? (Do the types of activities in the "resource sharing" course make the difference?). Defending these types of ascriptions to a program is prerequisite

[12]Defending causal claims is a thorny issue. For example, what if a person has a car accident because while driving he is stunned by a radio announcement that his company has just closed. This closure is due to a U.S. Presidential decision to ban imports of the company's principal product. Has President Reagan caused the car accident? (President Reagan's action set off the chain of events and the accident presumably would not have happened had Reagan not done so).

[13]See footnote dealing with Table 12: Cross-tabulation.

to assessing the merit of these outcomes.

ii. Assessment

Assessment refers to the normative judgements made from the data. The discussion of descriptive and evaluative categories in the previous section on tabulation raised the following question: "Does achievement of 90% on this test demonstrate ***excellent*** command of third world trade issues?". Perhaps the test was very simple or solely measured recall of facts. In these cases, scoring even 100% wouldn't demonstrate excellence in understanding. Or perhaps, the test is particularly demanding and a score of 80% or less may reflect high degree of understanding. Notice this is not the same as determining the validity of the standards discussed in the design phase. We have, presumably, accepted scores on this test as a relevant criteria for determining understanding of key issues. What we lack are ***norms*** to establish what a particular level of performance on this test signifies. In short, and this is the important point, the grounds for interpreting data in terms of the ***achievement*** of some standard must be identified and justified.

Setting norms

In the case of interpreting test scores, levels of performance (norms) may be set by a group of authorities in development issues and educational measurement. The putative merit of standardized tests is that they come with predetermined, justified norms. Alternatively, the test could be given to persons who are known to have excellent command (or poor command) of development issues. The scores obtained from these persons could be used to set norms. It is particularly difficult to set norms by which to interpret the significance of any change. Education is a slow process, met with delays and advanced largely by incremental improvements. Because of this, even modest amounts of change which can be attributed to a program are encouraging. Even small improvement, especially if the change is still in evidence months after the program, may be grounds for a positive assessment.

Assessment involves making judgements about each set of data in terms of meeting or failing to meet an evaluation standard. It will be remembered that, in the design stage, the use of multiple sources of evidence was recommended for each standard. Ideally, this means that the evaluator make several judgements about the achievement of a particular standard. If each of the pieces of evidence gives rise to the same assessment, then there are good grounds to conclude about the achievement/lack of achievement of that standard. If the evidence is contradictory, unless some of the evidence can be dismissed, then the evaluation is inconclusive on that standard. The same process applies when concluding about the measure of success in terms of the overall evaluation aim. Each interpretation of the degree of success on each standard is used to judge the overall success. Invariably, the verdict about the success/failure of the general aim is a qualified judgement. Rarely are all of the standards uniformly met or missed. In fact, in many cases, evaluators avoid making an overall assessment and limit the judgement to degrees of success/failure on each of the evaluation standards.

STAGE 9. REPORTING

There is no fixed format for reporting results to stakeholders. An oral presentation accompanied by a few duplicated handouts might be an ideal medium for a small group of animateurs. If the evaluation is detailed and the results are such that the stakeholders will likely want to study or at least refer to them a written report will be required. The overriding concern is how best to communicate (i.e., relate and explain) the results to the intended audience. This means reporting should be clear and candid without being brutal. It need not be voluminous nor exhaustive. There is no need to provide all the raw data nor, for that matter, to comment on all aspects of the evaluation. Rather the data should be reported only if it relates to the judgements and recommendations that the evaluators wish to make and defend.[14]

A formal evaluation report is often structured more or less as indicated in Table 13.

TABLE 13: SAMPLE OUTLINE – FORMAL REPORT

1. The Program
    - 1.1 Background/History
    - 1.2 Major Components
2. The Evaluation Study
    - 2.1 Purpose
    - 2.2 Evaluation Team
    - 2.3 Design and Methodology
    - 2.4 Data Collection and Analysis
    - 2.5 Problems/Caveats
3. The Findings
    - 3.1 Discussion of Aim or General Concern #1
    - 3.2 Discussion of Aim or General Concern #2
    - 3.3 Discussion of Aim or General Concern #3
4. Conclusions/Suggestions (if any)
5. Appendicies
    - 5.1 Extracts of key documents
    - 5.2 Tables/charts/maps
    - 5.3 Instruments
    - 5.4 Lists of respondents/subjects
    - 5.5 Bibliography of materials, articles, books

[14]This is not always the case. Sometimes it is difficult to decide what to make of certain data. Rather than defend a particular interpretation, the evaluator may prefer to report the data and leave the stakeholders to draw their own conclusions.

If an abridged report is made available for distribution to specific sub-groups the format may be altered as reflected in Table 14.

TABLE 14. SAMPLE OUTLINE – ABRIDGED REPORT

1. Strengths of project
2. Concerns about project
3. Suggestions

It is standard practice that the key stakeholders be given the opportunity to preview a draft of the major evaluation conclusions prior to finalization and broader distribution. Distribution of the report is normally at the discretion of the stakeholders.

STAGE 10. UTILIZATION OF RESULTS

What can be done to increase the likelihood that evaluation results will be used? Perhaps three of the most significant factors were addressed previously:

- set relevant aims for the evaluation
- involve key persons in all stages
- conduct a quality study

In the first stages of evaluation emphasis was placed on inclusion of stakeholders in the setting of aims. Obviously, if evaluators focus on actual concerns of involved parties, then there is increased likelihood of utility of results. Similarly, if persons who are in positions to assess and implement results are involved throughout the evaluation, then greater interest in results is to be expected. This implies including internal planners, managers, developers and animateurs. Their involvement should reduce the possiblity of last minute surprises and may increase the acceptance of the results. It may also lead to preliminary testing of possible solutions within the evaluation proper. Finally, there is no point in implementing results of a poorly conducted evaluation. More positively, a thorough evaluation is more likely to accurately identify and explicate the areas of concern. This means better information upon which to base remedial strategies.

The last comment brings us to the central issue in evaluation utilization. Strategies must be adopted which react to identified concerns and strengths. Sometimes recommendations or suggestions are included in the evaluation report. As likely, an action plan will have to be designed by parties involved with the program. The mechanisms for program planning are complex, however one aspect will be mentioned.

- arrange for follow-up

Follow-up provisions should be made to check that the strategies are in place and are having the desired effects. This may involve building into the program on-going monitoring of key indices or scheduling periodic verifications. Either way, persons charged with implementation should be made to feel their efforts to improve the program will be recognized.

SELECTED REFERENCES

There is no shortage of literature on development education and program evaluation. The problem is one of finding an entry point to this maze. The following selected references may serve as an adequate beginning.

1. Definitions

Theme Issue. *Canadian and International Education*, V.12, N.3, 1983.

United Nations Development Education Directory: For Non-Governmental Organizations in Industrialized Countries. Third edition, January 1983.

Stone, Tim. *A Whole New World: Development Education in Canada.* Ottawa: Canadian Council for International Co-operation, 1982.

Churchill, Stacy. "International Education and Education on Human Rights: The Canadian Experience". *Canadian and International Education*, V. 9, N. 1, 1980, 6-32.

Hanvey, Robert G. *An Attainable Global Perspective.* New York: Global Perspectives in Education, Inc., 1979.

Theme Issue. *The History and Social Science Teacher*, V. 13, N.4, Summer 1978.

Theme Issue. *Social Education*, V.41, N.1, January 1977.

Remy, Richard C., James A. Nathan, James M. Becker, and Judith V. Torney. *International Learning and International Education in a Global Age.* Washington, D.C: NCSS, 1975.

2. Critiques

Popkewitz, Thomas S. "Global Education as a Slogan System". *Curriculum Inquiry*, V. 10, N. 3, 1980, 303-316.

Daniels, L.B. "Development Education: The 20th Century Survival and Fulfillment Skill". *Canadian and International Education*, V.12 N.3, 1983, 199-204.

Perras, Jean and Francine Trempe. "Study of International Development Education in Canada". Final Report. Montreal: Canadian Council for Internation Co-operation, May 1982.

3. Evaluation

Pietro, Daniel Santo (ed.). *Evaluation Sourcebook For Private and Voluntary Organizations.* New York: American Council of Voluntary Agencies for

Foreign Service, Inc., 1983.

Sudman, Seymour and Bradburn, Norman N. *Asking Questions: A Practical Guide to Questionnaire Design.* San Fransisco: Jossey-Bass Publishers, 1983.

Hamilton, Stephen. "Evaluating Your Own Program". *Educational Leadership*, V. 37, N. 7, April 1980, 545-552.

Werner, Walter. "Editorial Criticism in Curricular Analysis". *Curriculum Inquiry*, V. 10, N. 2, 1980, 143-154.